Census 2000 Evaluation O.10
April 8, 2003

Housing Unit Duplication in Census 2000

FINAL REPORT

This evaluation study reports the results of research and analysis undertaken by the U.S. Census Bureau. It is part of a broad program, the Census 2000 Testing, Experimentation, and Evaluation (TXE) Program, designed to assess Census 2000 and to inform 2010 Census planning. Findings from the Census 2000 TXE Program reports are integrated into topic reports that provide context and background for broader interpretation of results.

John Jones
Decennial Statistical Studies Division

U S C E N S U S B U R E A U
Helping You Make Informed Decisions

CONTENTS

LIST OF TABLES

EXECUTIVE SUMMARY

This paper examines housing unit duplication in Census 2000 as measured by the 2000 Accuracy and Coverage Evaluation. The Accuracy and Coverage Evaluation was an operation undertaken to evaluate the population and housing coverage of Census 2000. First, it performed an independent enumeration of housing units and people within a stratified sample of census block clusters. Then it matched this enumeration against the Census 2000 enumeration of housing units and people in those same block clusters. The Accuracy and Coverage Evaluation included an initial housing unit phase, where housing units in the sampled block clusters were matched against units listed in the January 2000 Decennial Master Address File in those same clusters; a person interview phase, during which demographic information was collected from census day residents of housing units in the sampled block clusters; a person match phase, during which persons listed in the independent enumeration were matched against the census record of persons in those same clusters; and a final housing unit phase, during which updates to housing unit inventories after the end of the initial housing unit phase were processed. Estimates of person and housing unit coverage were produced after the completion of the Accuracy and Coverage Evaluation.

To produce housing unit coverage estimates, the Accuracy and Coverage Evaluation classified census housing units as either correct enumerations or erroneous enumerations. Correctly enumerated housing units were confirmed to exist as housing units within the block cluster on census day while erroneously enumerated housing units were not confirmed to exist within the block cluster as housing units on census day. Because duplicate housing units were erroneous enumerations, the initial and final housing unit phases of the Accuracy and Coverage Evaluation. conducted a search for housing unit duplicates and identified them as such. The objective of this study is to document the extent of census housing duplication, to give the characteristics of housing units most likely to be duplicates, and to identify the nature of duplicate housing unit addresses.

The following limitations apply:

- **Reinstated housing units are not included.** Housing units that were reinstated in the census as a result of the census operation to remove housing unit duplicates are outside of the scope of this study.

- **The duplicate search area was limited.** The Accuracy and Coverage Evaluation search area was primarily limited to the sampled block clusters. In block clusters chosen for extended search, the search area was extended to one ring of blocks surrounding the block cluster. Consequently, duplicates of units located outside of the search area were not counted.

The key findings of this study are as follows:

- **Mailout/mailback areas had significantly less housing unit duplication than other enumeration areas.** Mailout/mailback areas consist of the most populous areas and contain city-style addresses while other enumeration areas consist of more isolated areas that often contain non-city-style addresses.

- **There was significantly more duplication among vacant housing units than among occupied housing units.** In particular, there is less duplication in homes occupied by homeowners.

- **Housing unit duplication was more prevalent in small multi-unit structures.** This is in comparison with single unit structures and with large multi-unit structures.

- **Housing unit addresses coded and classified to be duplicates of one another generally were not identical.** There was substantial disagreement on address characteristics.

1. BACKGROUND

The 2000 Accuracy and Coverage Evaluation (A.C.E) included the match of an independent enumeration of housing units in a sample of block clusters against the Census 2000 enumeration of housing units in those same clusters. The independent enumeration is known as the P-sample. After the initial housing unit phase, census housing units in the sample clusters were subsampled. Units remaining in sample after this subsampling belong to the E-sample. The A.C.E. determined the enumeration status of E-sample units only. Therefore, only E-sample units are within the scope of this study.

The housing unit phase began with an independent listing of the addresses of all of the housing units in the sample clusters. The sample clusters were stratified into small, medium, and large block clusters, and clusters located on American Indian reservations. After the independent listing, there was a reduction in the number of small block and medium block clusters. (This reduction was distinct from a subsequent subsampling of housing units within large block clusters). After this reduction in the number of sample clusters, housing units on the independent list of sample addresses were matched against the housing units listed on the January 2000 version of the Decennial Master Address File (DMAF) for sampled clusters.

The housing unit matching began with a computer match of addresses that compared the independent listings with the DMAF (census) listings and identified matched addresses and possibly matched addresses. Addresses were said to match when an address from the independent list and an address from the census referred to the same housing unit. All addresses not matched by computer then came under before followup (BFU) clerical review where additional matches were made and duplicate searches within census address listings were performed. Addresses that were nonmatched, possibly matched, or determined to be possible duplicates of other addresses after BFU clerical review were sent to housing unit followup (HUFU). It was possible that two addresses not identified as potential duplicates by the clerical review could be identified as duplicates by HUFU. Information obtained from housing unit followup was used to assign after followup (AFU) match and enumeration codes to housing units.

Duplicates occurred when two or more addresses referred to the same housing unit. When this happened, one of the addresses was regarded to be the primary (true) address, and the others were considered to be duplicate addresses. The primary address often was matched to an address on the independent list, or otherwise confirmed to be correctly enumerated in the block cluster. Duplicate linkages between the primary address and the duplicate address were generated. There were duplicate linkages between E-sample and non E-sample addresses, as well as linkages between E-sample addresses. Duplicate search also occurred in the final housing unit (FHU) phase. Some of the updates to census housing unit inventory created additional duplication and some of the units in relisted and list enumerate clusters were duplicates. Possible duplicate addresses were sent for confirmation during final housing unit followup (FHUFU).

2. METHODS

Table 1 gives the overall weighted percentage of E-sample housing units classified as duplicates while **Tables 2 through 9** give this percentage and the associated standard error for each level of the following variables:

- Region
- Sampling Stratum
- Metropolitan Statistical Area/Type of Enumeration Area (MSA/TEA) group
- Tenure
- Type of structure (Number of units at Basic Street Address)
- Tenure by Number of units at BSA
- MSA/TEA group by Number of Units at BSA
- Race/Hispanic origin of householder (Occupied units only)

The percentage of duplication is the ratio of the weighted number of duplicates to the weighted number of housing units multiplied by 100. Both units with final match code as duplicate and units with duplicate links to non E-sample housing units are counted as duplicates. Units with final match code as duplicate are counted as one erroneous enumeration while units with duplicate links to non E-sample housing units are counted as a partial erroneous enumeration, with the exact fraction depending upon the number of non E-sample duplicate links. Standard errors of the duplicate percentage were calculated using the stratified jackknife. For a given variable, each pair of levels was compared by a t-test with a critical value of t given below each table. The critical values are based on a multiple comparison of means technique with a Bonferroni adjustment. The overall significance level is 10 percent.

Tables 10 through 14 utilize a database of linked duplicate pairs. If an E-sample unit had *n* duplicates then the database had *n* separate records. Each record of the database contains address and housing unit characteristics of each member of the linked duplicate pair. The database was used to investigate the agreement on these characteristics of the linked pairs.

3. LIMITS

A major limitation of this study concerns the available universe of eligible housing units. Before the beginning of the 2000 Accuracy and Coverage Evaluation, the census flagged housing units it thought to be potential duplicates and removed these units from the existing housing unit inventory. The census flagged approximately 2.4 million housing units for this purpose. Some of these flagged units later were deleted (1.4 million) while some were kept in the census (1 million). Neither the units originally deleted nor the units subsequently reinstated were within the Accuracy and Coverage Evaluation universe. These flagged units are, therefore, outside of the scope of this study.

Other limitations include:

- The amount of duplication may be misstated because of coding restrictions. For example, there were no duplicate links allowed between duplicate units and other erroneously enumerated units.

- The duplicate search area was limited to the sample block clusters. In clusters chosen for extended search, the search area was the sample block clusters and the first ring of surrounding blocks. Therefore, a housing unit within the sample clusters that was the duplicate of a housing unit outside of the search area would not be included in this analysis. Furthermore, there are no available means of measuring the number of within search area housing units that are duplicates of outside of search area housing units.

- No data from Puerto Rico are included in this analysis.

4. RESULTS

4.1 What was the overall percent of census housing unit duplication in Census 2000 as measured by the 2000 Accuracy and Coverage Evaluation and how did it compare to that percent in the 1990 Census as measured by the 1990 Post-Enumeration Survey?

Table 1 gives percentages of E-sample housing unit duplication in Census 2000 and in the 1990 Census. It also gives an estimate of the percentage of housing unit duplication in the 1980 Census. The 2000 percentages are measured by the 2000 Accuracy and Coverage Evaluation while the 1990 percentage is measured by the 1990 Housing Unit Coverage Study (HUCS). The 1990 HUCS was based upon a sample of half of the housing units that were in the 1990 Post-Enumeration Survey (see Childers, 1993). The first column gives percentages of the weighted total of erroneously enumerated housing units that were duplicates. The second column gives the percentages of the weighted total of E-sample housing units that were coded as erroneous enumerations while the third column gives the percentages of the weighted total of E-sample housing units that were duplicates. The final column gives a nationwide estimate of the number of census housing units that were duplicates. Results in the final two columns show that housing unit duplication decreased in Census 2000, possibly as a result of the census housing unit duplicate operation.

Table 1: Overall Percent E-sample Housing Unit Duplication

Year	Percent of Erroneous Enumerations that were duplicates	Percent of E-sample housing units that were Erroneous Enumerations	Percent of E-sample housing units that were duplicates	Estimated number of census duplicates
1980	NA	NA	1.01*	947,084
1990	33.4	2.8	0.95	971,505
2000	24.8	2.3	0.57	660,656

*This is the result of a special study of the 1980 census and not of a comprehensive coverage measurement survey. For details see O'Brien. The 1980 figures are not directly comparable to the 1990 and 2000 figures. For details, see Childers (1993).

4.2 What was the frequency of census housing unit duplication? Did this frequency vary across different groupings of important variables?

Tables 2-8 give weighted percentages of census housing unit duplication in the 2000 Accuracy and Coverage Evaluation by important variables. They display variable level names, variable level numbers, the weighted percentage of E-sample housing units in level that were duplicates (percent duplicate), the stratified jackknife standard error (s.e.), a list of level numbers with which a significant difference was found (differ from), and the weighted percent of E-sample housing units belonging to each variable level (Percent E sample). For a given variable, each pair of levels of each variable was compared by a t-test with a critical value that reflects the Bonferroni criterion. These critical values of t are given below each table.

Table 2 gives weighted housing unit duplication rates, by region. It shows that there were no significant regional differences in housing unit duplication.

Table 2: Weighted E-sample Housing Unit Duplication Percentages, by Region

Region	Percent Duplicate	(s.e.)	Rank	Differ from	Percent E sample
Northeast	0.68	(0.12)	2	none	19.2
Midwest	0.39	(0.06)	4	none	23.0
South	0.71	(0.19)	1	none	36.8
West	0.43	(0.08)	3	none	21.0

Critical value of t: 2.386

During the initial housing unit phase, sample clusters were stratified into small block clusters (less than three housing units per block), medium block clusters (from 3-79 housing units per block), large block clusters (80 or more housing units per block), and clusters located on American Indian reservations (A.I.R.). **Table 3** gives weighted housing unit duplication rates, by sampling stratum. Results show that clusters on American Indian reservations have a significantly higher percentage of housing unit duplication than medium sized clusters. There were no other significant differences.

Table 3: Weighted E-sample Housing Unit Duplication Percentages, by Sampling Stratum

Stratum	Percent Duplicate	(s.e.)	Rank	Differ from	Percent E sample
Small	0.64	(0.29)	3	none	0.6
Medium	0.48	(0.03)	4	1	60.7
Large	0.72	(0.20)	2	none	38.4
A.I.R.	1.50	(0.38)	1	4	0.3

Critical value of t: 2.386

Table 4 gives housing unit duplication rates, by Metropolitan Statistical Area / Type of Enumeration Area (MSA/TEA) Group. Mailout/mailback areas generally have city-style addresses and most of these addresses lie in large or medium metropolitan statistical areas (MSA). Other types of enumeration areas consist of isolated areas with small populations and non-city-style addresses. There was a significantly higher percentage of housing unit duplication in the more rural isolated areas.

Table 4: Weighted E-sample Housing Unit Duplication Percentages, by Metropolitan Statistical Area/ Type of Enumeration Area (MSATEA)

MSA/TEA	Percent Duplicate	(s.e.)	Rank	Differ from	Percent E sample
Large MSA Mailout/Mailback	0.31	(0.04)	4	1	28.0
Medium MSA Mailout/Mailback	0.35	(0.07)	3	1	31.0
Small MSA & Non MSA Mailout/Mailback	0.83	(0.33)	2	none	21.4
All Other TEAs	1.01	(0.08)	1	3,4	19.6

Critical value of t: 2.386

Table 5 gives housing unit duplication frequency by the type of housing unit structure. Small multi-unit structures (from two to nine units at basic street address, inclusive) had a higher percentage of duplicates than single unit structures and the difference is significant.

Table 5: E-sample Housing Unit Duplication Percentages by Type of Structure

Number of units at address	Percent Duplicate	(s.e.)	Rank	Differ from	Percent E sample
1	0.36	(0.03)	3	1	72.9
2-9	1.40	(0.44)	1	3	11.6
10+	0.96	(0.77)	2	none	15.5

Critical value of t: 2.121

Table 6 gives housing unit duplication frequencies by MSA/TEA cross classified with Type of Structure. Small multi-unit structures have higher duplication frequency percentages in all areas, but the percentage is especially high in the non Mailout/ Mailback areas.

Table 6: E-sample Housing Unit Duplication Percentages, by MSA/TEA cross classified with Type of Structure

Variable level	Percent Duplicate	(s.e.)	Rank	Differ from	Percent E sample
Large MSATEA single unit	0.13	0.03	12	1,3,4.6,7	16.62
Large MSATEA 2-9 units	1.00	0.13	6	1,9,10,11,12	4.65
Large MSATEA 10+ units	0.30	0.10	10	1,3,4,6,7,11	6.74
Medium MSATEA single unit	0.17	0.03	11	3,4,6,10	22.40
Medium MSATEA 2-9 units	1.15	0.19	4	1,9,10,11,12	3.53
Medium MSATEA 10+ units	0.61	0.39	8	1	5.11
Small MSATEA single unit	0.36	0.08	9	1,3,4,6,7	16.13
Small MSATEA 2-9 units	1.34	0.27	3	1,9,10,11,12	2.25
Small MSATEA 10+ units	2.98	1.87	2	none	3.02
Other MSATEA single unit	0.82	0.06	7	1,9,10,12	17.79
Other MSATEA 2-9 units	3.96	0.66	1	3,4,5,6,7,8,9,10,12	1.11
Other MSATEA 10+ units	1.12	0.60	5	1	0.64

Critical value of t: 3.164

Table 7 gives housing unit duplication percentages by housing unit tenure. It shows that vacant units have the highest percentage of duplicates and that owner occupied housing units have a significantly lower housing unit duplication percentage than units not occupied by the owner.

Table 7: E-sample Housing Unit Duplication Percentage, by Housing Tenure

Tenure	Percent Duplicate	(s.e.)	Rank	Differ from	Percent E sample
Owner	0.34	(0.08)	3	1,2	60.8
Non Owner	0.62	(0.03)	2	3	30.3
Vacant	2.01	(0.70)	1	3	8.9

Critical value of t: 2.121

Table 8 gives housing unit duplication percentages by housing tenure cross classified with the type of structure variable. It shows that the percentage of duplicates was highest among vacant units, particularly vacant multiunits. Owner occupied small multi-unit structures had a high percentage of duplicates.

Table 8: E-sample Housing Unit Duplication Percentages by Housing Tenure cross classified with Type of Structure

Variable level	Percent Duplicate	(s.e.)	Rank	Differ from	Percent E sample
Owner single	0.24	0.02	9	2,3,4,5,8	55.89
Owner 2-9	2.07	0.25	3	2,5,7,8,9	2.38
Owner 10+	0.84	0.40	6	none	2.49
Nonowner single	0.43	0.05	8	2,3,4,5,9	11.25
Nonowner 2-9	1.00	0.10	5	2,3,9	7.86
Nonowner 10+	0.56	0.18	7	2,3	11.19
Vacant single	1.38	0.21	4	8,9	5.80
Vacant 2-9	2.58	0.43	2	3,5,7,8,9	1.31
Vacant 10+	3.58	2.81	1	none	1.82

Critical value of t: 3.051

Table 9 gives housing unit duplication percentages by race domain of householder, for occupied units only. It shows that there were no significant racial differences in the frequency of housing unit duplication.

Table 9: Weighted E-sample Housing Unit Duplication Percentages, by Racial/Ethnic Domain (Occupied Units Only)

Domain	Percent Duplicate	(s.e.)	Rank	Differ from	Percent E sample
American Indian on reservation	1.31	(0.38)	1	none	0.1
American Indian off reservation	0.61	(0.21)	2	none	0.5
Hispanic	0.58	(0.09)	3	none	8.7
Non-Hispanic black	0.53	(0.06)	4	none	11.3
Native Hawaiian or Pacific Islander	0.29	(0.15)	7	none	0.2
Non-Hispanic Asian	0.48	(0.13)	5	none	3.0
Non-Hispanic white	0.40	(0.04)	6	none	76.2

Critical Value of t: 2.815

4.3 What attributes did housing units coded as duplicate have in common with those that they duplicated? On which attributes did duplicate housing unit pairs disagree?

Housing units coded as duplicates have been linked with their corresponding primary and the address characteristics of each have been compared. **Table 10** gives the unweighted percentage of primary duplicate pairs that agree on each of six different address characteristics. When both of the linked addresses was missing a characteristic, the pair was not included in the percentage calculation for that characteristic. The final column of Table 10 gives the percentage of the total number of linked duplicate pairs that were included in the percentage agreement calculation. Results show that for duplicate pairs there was widespread disagreement on the rural route and unit designation characteristics, extensive disagreement on house number and street name, but that there was relatively less disagreement on ZIP Code and census block.

Table 10: Percentage Agreement on Address Characteristics of E-sample Primary-Duplicate Pairs

Address Characteristic	Percent Agree	Percent Disagree	Percent included
Rural Route and Box Number	12.6	87.4	13.6
Unit Designation	10.9	89.1	49.2
House Number	50.2	49.8	85.7
Street Name	54.9	45.1	88.2
ZIP Code	78.2	21.8	96.4
Census Block	78.5	21.5	100.0

There was extensive agreement on ZIP Code, and census block. Disagreement on ZIP Code and census block suggests that the duplicate housing unit may have been incorrectly geocoded to the census block. There was more extensive disagreement on rural route, unit designation, house number, and street name. For street name and unit designation, there were different ways in which the disagreement between primary and duplicate occurred. Disagreement in street name basically occurred in one of the following ways:

- **Missing street name**: One of the linked pair of addresses had no street name.
- **Different street name**: The linked pair had completely different street names.
- **Same street name but different street identifiers**: Here street identifiers tell what type of street the unit is on. Examples include ave (avenue), pl (place), blvd (boulevard).

Identifiers also include direction indicators such as N (north) or SW (southwest). Pairs in which one address had an identifier and the other did not would be under this heading.

- **Different spelling of same street name**: This also applies when one address had the street name in one word while the other has it in two words.

Table 11 gives the percentage of street names that disagreed that are in each of the aforementioned categories:

Table 11: Types of Disagreeing Street Names

Type	Percent
Missing Street Name	39.4
Different Street Name	23.5
Different Street Identifiers	24.1
Different Spelling of Same Street Name	13.0

Disagreement in unit designation occurred in one of the following ways:

- **Missing Unit Designation**: One of the linked pair of addresses had no unit designation.
- **Different Unit Designation**: This refers to categorically different designations, such as a linked pair with designations BASEMT and APT 2.
- **Different Unit Numbers**: This refers to different numbers or letters.
- **Same Unit Number and Different Identifier**: Here identifiers refer to the words or symbols in a designation. Examples include APT 4 or # 5.

Table 12 gives the percent of linked pairs with disagreeing unit designations that are in each of the aforementioned categories.

Table 12: Types of Disagreeing Unit Designations

Type	Percent
Missing Unit Designation	49.2
Different Unit Designations	15.0
Different Unit Numbers	11.0
Same Unit Number and Different Identifier	24.8

Table 13 gives the percentage agreement on the six address characteristics by type of enumeration area. The first three levels of the MSA/TEA variable of Table 4 were collapsed to form the mailout/mailback level. Results show that there was relatively more disagreement in the non mailout/mailback areas, particularly in house number and street name.

Table 13: Percentage Agreement of E-sample Linked Pairs on Address Characteristics, by TEA

Address Characteristic	Mailout/Mailback		Non Mailout/Mailback	
	Agree	Disagree	Agree	Disagree
Rural Route & Box Number	0.0	100.0*	12.9	87.1
Unit Designation	10.8	89.2	15.6	84.4
House number	64.9	35.1	18.6	81.4
Street name	68.9	31.1	26.4	73.6
ZIP Code	87.7	12.3	61.2	38.8
Census Block	73.3	26.7	86.1	13.9

* There were only 5 linked pairs falling into this category

Table 14 gives the percentage agreement on all six address characteristics by type of structure. The last two levels of the type of structure variable of Table 5 were collapsed to form the multi-unit level. It shows that there was relatively less disagreement on the house number, street name, unit designation, and rural route characteristics in multi-units compared to single units.

Table 14: Percentage Agreement of E-sample Linked Pairs on Address Characteristics, by Type of Structure

Address Characteristic	Single Unit		Multi-unit	
	Agree	Disagree	Agree	Disagree
Rural Route & Box Number	12.4	87.6	13.5	86.5
Unit Designation	4.3	95.7	12.3	87.7
House number	30.2	69.8	66.8	33.2
Street name	32.6	67.4	74.0	26.0
ZIP Code	69.1	30.9	85.4	14.6
Census Block	78.1	21.9	78.9	21.1

5. CONCLUSIONS AND RECOMMENDATIONS

The objective of this study was to document the extent of census housing duplication, to identify the characteristics of housing units most likely to be duplicates, and to identify how duplicate housing unit addresses compare to one another. Results of this study can be used to identify characteristics and geographic areas that may be most beneficial to study or target when searching for duplicates. The results can be used to guide unduplication efforts and to help correct erroneous addresses.

The major conclusions are as follows:

There was substantial duplication of Census 2000 housing units. Table 1 shows that nearly 25 percent of all erroneously enumerated housing units were duplicates in 2000. This figure is lower than the corresponding figure of 33.4 percent in 1990, but it still is significant. It shows that it is beneficial to conduct duplicate housing unit searches, and that successful efforts to unduplicate housing units can result in better housing unit coverage estimates.

Housing unit duplication was not uniform. It varied, sometimes rather widely, by size of urban area, by whether units were single units or multiunits, by the occupancy status of the housing unit, and by the housing tenure of the occupant. Housing unit duplication was more frequent in relatively small segments of the E-sample.

The next major conclusions relate to the location and kind of housing units that were most likely to have duplicates:

There was more housing unit duplication in small cities and in rural areas.
Table 4 shows that the percentage of housing unit duplication increases as you proceed from large mailout/mailback areas to small mailout/mailback areas and non mailout/mailback areas. The highest duplication percentages were concentrated in the 41 percent of the E-sample containing the small mailout/mailback and the 'Other TEA' areas. The percentage in non mailout/mailback areas is the largest (1.01 percent), and these are sparsely populated and geographically isolated areas where census information is given to enumerators. The results imply that duplicate search and unduplication efforts should be targeted to small cities and rural areas.

There was more housing unit duplication among units in small multi-unit structures than among single unit structures. In particular, Table 5 shows that the difference was significant. highest in the small multi-unit structures that had between 2 and 9 units at a basic street address. Table 6 shows extensive duplication among units in small multi-unit structures in large, medium and small cities and in isolated non mailout/mailback (Other MSATEA) areas. For each of these areas, the duplication percentage is significantly higher in small multi-units than in single units. The results suggest that duplicate search and unduplication efforts be targeted to multi-unit structures.

There was more housing unit duplication among vacant units than among occupied units. Table 7 shows that the vacants had the highest percentage of duplication while taking approximately 9 percent of the E-sample. Table 8 shows that three (1.38, 2.58, and 3.58 percent) of the four highest observed duplication percentages are among vacant units, although they are not significantly different from all of the other types of units. In conjunction with the table contrasting single and multiunits, Table 8 implies that it is most beneficial to perform duplicate search on and to unduplicate single units when they are vacant.

American Indian reservations have a relatively higher housing unit duplication percentage. Table 3 shows that block clusters located on these reservations had the highest duplication percentage, while taking less than 1 percent of the total number of E-sample housing units. Table 9 shows that, although not significantly different, householders that are American Indians on reservations had the highest observed duplication percentage among occupied units (1.31 percent). These results suggest that duplicate search and unduplication operations should occur on reservations.

The final conclusion concerns the nature of duplicate housing unit addresses:

Duplicate addresses that referred to the same housing unit seldom were identical. Perhaps this is why they were not detected as possible duplicates by the census duplicate housing unit operation (see section on Limits). Non-city-style (rural route and box number) addresses and unit designations seldom agreed. The percentages were12.6 percent and 10.9 percent, respectively. House number and street name agreed about 50 percent of the time, while ZIP Codes and census blocks agreed most often. Both were around 78 percent. When addresses with duplicate links disagreed on streetname or unit designation, it often was because one of the addresses had a missing address characteristic. In general, mailout/mailback areas had less disagreement than non mailout/mailback areas, which suggests that there is potential for address improvement in these areas.

6. REFERENCES

Barrett, D., Beaghen, M., Smith, D., Burcham, J. (2001) "ESCAP II: Census 2000 Housing Unit Coverage Study", Executive Steering Committee for A.C.E. Policy (ESCAP II), Report 17, October 17, 2001, U.S. Census Bureau

Childers, D. (2001). "The Design of the Census 2000 Accuracy and Coverage Evaluation (A.C.E.)" DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-01, January 11, 2000, U.S. Census Bureau

Childers, D. (1993). "Coverage of Housing in 1990 Decennial Census" 1990 Decennial Census Preliminary Research and Evaluation Memo No. 253, October 7, 1993, U.S. Census Bureau

Childers, D. (1992). "The 1990 Housing Unit Coverage Study," *Proceedings of the Section on Survey Research Methods,* American Statistical Association, 506-11

Fay, R. (1990). "VPLX: Variance Estimates for Complex Samples," *Proceedings of the Section on Survey Research Methods,* American Statistical Association, 266-71

Hocking, RR (1986). *Methods and Applications of LinearModels: Regression and the Analysis of Variance* (New York: John Wiley and sons), pp 108-9

Nash, Fay (2000), "Results of Reinstatement Rules for the Housing Unit Duplication Operations," internal document, November 21, 2000, U.S. Census Bureau.

O'Brien, James L. (1983) . "Overcount Evaluation; Housing Unit Enumeration Duplication Study–Results" Preliminary Evaluation Results Memorandum No.44, February 28, 1983, U.S. Census Bureau.